INFORMATION AND CONTROL 1, 233–251 (1969)

The Capacity of the General Time-Discrete Channel with Finite Alphabet

JACOB ZIV*

Scientific Department, Israel Ministry of Defence P.O.B. 7063, Tel-Aviv, Israel

In this paper, the capacity, C, of the general discrete time-varying channel with memory is derived. It is also demonstrated that for rates below some positive number C_s, the upper bound on the probability of of error decreases exponentially with the block length, n. (Clearly $C_s \leqq C$). The capacity C is derived by proving a coding theorem which is then followed by a (weak) converse theorem.

1. INTRODUCTION

Let us consider a set A of L letters, which we call the "input alphabet" and a set B of M letters which we call the "output alphabet".

Let us consider sequences of letters infinite on both sides

$$x = \cdots x_{-1}, x_0, x_1 \cdots ; \quad x_i \in A$$

$$y = \cdots y_{-1}, y_0, y_1 \cdots ; \quad y_i \in B$$

The sequence x will be regarded as an elementary event in a probability space A^1 and the sequence y will be regarded as an elementary event in a probability space B^1.

Consider now the following cylinder sets (n-sequences):

a) The set X_n^j of all elementary events x such that $x_i = a_i$; $j \leqq i \leqq j + n - 1$; a_i is some element of A

b) The set Y_n^j of all elementary events y such that $y_i = b_i$; $j \leqq i \leqq j + n - 1$; b_i is some element of B

The general discrete channel is characterized by a system of conditional probabilities

$$P(Y_n^j \mid X_n^j, s^j)$$

$$j = 0 \pm 1, 2 \cdots ; \quad n = 1, 2 \cdots ; \quad s^j \in S^j$$

* On leave at Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey 07974.

where S^j is a countable set of all possible channel states at the jth instance. Let $P(s^j)$ be a probability distribution defined on S^j.

The channel is assumed to be "non-anticipative": i.e. $Y_n{}^j$ does not depend on future inputs.

A code (N, n, j) is a set of N input n-sequences $X^j_{n,1}$, $X^j_{n,2} \cdots X^j_{n,N}$, together with a corresponding decision function which is assigning to each output n-sequence $Y_n{}^j$, one of the input sequences

$$X^j_{n,1} \cdots X^j_{n,N}$$

A code $(N, n, \lambda^j_{n,s})$ is a code consisting of N input sequences of n letters, whose maximum probability of error that is associated with a state $s^j \in S^j$ is $\lambda^j_{n,s}$, assuming that both the sender and receiver do not know the state of the channel s^j. In addition, let

$$\lambda_n{}^j = \sum_{S^j} p(S^j)\lambda^j_{n,s}$$

be called the average probability of error.

A code $(N, n, \lambda_n{}^j)$ is a code consisting of N input sequences of n letters whose probability of error is $\lambda_n{}^j$. (Ash, 1965; Wolfowitz, 1964)

Let the information rate R be defined as

$$R = \frac{1}{n} \log N$$

we then prove the following theorems:

THE CODING THEOREM. (a) *For any R smaller than some number C which is nonnegative, and for any $j = 0, \pm 1, \pm 2, \cdots$, there exist codes $([e^{nR}]n, \lambda_n{}^j)$ such that $\lim_{n\to\infty} \lambda_n{}^j = 0$. The number C is determined by the system of probabilities $P(S^j)$ and $P(Y_n{}^j \mid X_n{}^j, S^j)$.*

(b) *For any R smaller than some nonnegative number C_s, there exist codes $([e^{nR}], n, \lambda_n{}^j)$ such that for any $j = 0, \pm 1, \pm 2, \cdots$*

$$\lambda_n{}^j \leqq e^{-E(R)\cdot n} \quad \textit{for large enough } n,$$

where $E(R)$ is positive and is independent of n. The number C_s is determined by the system of conditional probabilities $P(Y_n{}^j \mid X_n{}^j, S^j)$.

(c) *For any R smaller than some nonnegative number C_s there exist codes $([e^{nR}, n, \lambda^j_{n,s})$ such that for large enough n, any $j = 0, \pm 1, \pm 2, \cdots$ and any s^j, $\lambda^j_{n,s} \leqq K(s^j)e^{-E(R)n}$ where $K(s^j)$ is independent of n and R. Thus, $\sup_{s^j} \lim_{n\to\infty} \lambda^j_{n,s} = 0$.*

THE CONVERSE THEOREM. (a) *Given any sequence of codes $([e^{nR}], n,$*

$\lambda_n{}^j$) *then if* $R > C$, *there always exists a positive number* λ *such that for sufficiently large* n_0, $\lambda_n{}^j > \lambda$ *for some* $n > n_0$ *and some* j.

b) *Given any sequence of codes* ($[e^{nR}]$, n, $\lambda_{n,s}^j$) *then if* $R > C_s$ *there always exists a positive number* λ *such that for sufficiently large* n_0, $\sup_{s^j} \lambda_{n,s}^j > \lambda$ *for some* $n > n_0$ *and some* j.

The number C is called "channel capacity" and is defined as follows: Let

$$I_s{}^j(n) = \frac{1}{n} \sum_{\{X_n{}^j,\, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid s^j) \log \frac{P(Y_n{}^j \mid X_n{}^j, s^j)}{P(Y_n{}^j \mid s^j)} \tag{1}$$

and let

$$Q^0 = \{q^0 : \lim_{n\to\infty} \sup_K P\{s^j : I_s{}^j(n) < q^0\} = 0\} \tag{2}$$

where $P\{s^j : I_s{}^j(n) < q^0\}$ is the probability measure of all states s^j such that $I_s{}^j(n) < q^0$ and where

$$j = j_0 \pm Kn; \quad K = 0, 1, 2, 3, \cdots$$

Then

$$C = \inf_{j_0} \sup_{\mu} \sup_{q^0 \in Q^0} \{q^0\} \tag{3}$$

where μ is the probability measure that is defined on A^1. Clearly C is determined by the system of probabilities $P(s^j)$ and $P(Y_n{}^j \mid X_n{}^j, s^j)$.

The number C_s is given by:

$$C_s = \inf_{j_0} \sup_{\mu} \liminf_{n\to\infty} \inf_{K} \inf_{s^j} I_s{}^j(n) \qquad j = j_0 \pm Kn; \quad K = 0, 1, 2, 3, \cdots \tag{4}$$

Discussion

For any $R < C$, the average probability of error, $\lambda_n{}^j$, vanishes as n gets large. On the other hand when $R < C_s$, $\lambda_n{}^j$ as well as $\lambda_{n,s}^j$ decreases at least exponentially with n. Furthermore, $\sup_{s^j} \lambda_{n,s}^j$ cannot be made arbitrarily small once R is *larger* than C_s. Thus, if the set S^j is *finite* (a finite-state time varying channel) and if states with zero probability measure are to be ignored, C_s must be equal to C.

In this case, $\lambda_n{}^j$ (and $\lambda_{n,s}^j$) decreases at least exponentially with n for any rate up to channel capacity. A similar result has been obtained by Yudkin for finite-state stationary channels (Yudkin, 1967).

If, on the other hand, the set S^j is infinite, it is possible to find channels

for which $C_s < C$. Some of these channels may have the property that for rates in the region $C_s < R < C$ the probability of error decreases with n at a nonexponential rate.

We shall now bring an example for which $C_s = C$ where S^j is an infinite (countable) set, and also derive the capacity of Wolfowitz's compound channel in the special case where S^j is finite.

(1) *A Channel with Vanishing Memory* (c.v.m.)

Let a channel with vanishing memory be a channel for which for any probability measure μ on A^1,

$$\lim_{n\to\infty} | I^j_{s_K}(n) - I^j_{s_e}(n) | = 0 \tag{5}$$

uniformally with respect to j, where $s_K{}^j$ and $s_e{}^j$ are any two elements of S^j.

It follows that for any c.v.m. we have

$$C_{\text{c.v.m.}} = C_S = C = \inf_{j_0} \sup_{\mu} \lim_{n\to\infty} \inf_{K} I^j(n)$$

$$j = j_0 \pm Kn\,; \quad K = 0, 1, 2, 3. \tag{6}$$

where

$$I^j(n) \triangleq \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j) \log \frac{P(Y_n{}^j \mid X_n{}^j)}{P(Y_n{}^j)}. \tag{7}$$

In the special case where the channel is *stationary* and *memoryless* (d.m.c.)

$$P(Y_n{}^j \mid X_n{}^j, s^j) = \sum_{i=1}^{n} p(y_i \mid x_i) \qquad x_i \in A, \quad y_i \in B$$

where the conditional probability distribution $p(\cdot/\cdot)$ which is independent of i and of the channel state $s^j \in S^j$.
Thus

$$C_{\text{d.m.c.}} = \sup_{\mu_1(\cdot)} \sum_{\{x_1, y_1\}} \mu_1(x_1) P(y_1 \mid x_1) \log \frac{P(y_1 \mid x_1)}{P(y_1)} \tag{8}$$

where $\mu_1(\cdot)$ is the probability distribution on A (Wolfowitz, 1964).

(2) *The Compound Channel*

Let us consider the special case where the channel is *stationary* and is characterized by a set of conditional probability distributions, as

follows:

$$P(Y_n^j \mid X_n^j, s^j) = P(Y_n \mid X_n, s^j) = \prod_{i=1}^{n} p(y_i \mid x_i, s^j)$$

where the conditional probability distribution $p(\cdot \mid \cdot, s^j)$ for a given state s^j is independent of i; and $s^j \in S$ where S is a *finite* set. ($x_i \in A$; $y_i \in B$). Then, in this special case, we have by the coding and converse theorems of this paper that the capacity of this compound channel is:

$$C_s = C_{\text{compound}} = \sup_{\mu_1(\cdot)} \min_{s^j} \sum_{\{x_1, y_1\}} \mu_1(x_1) P(y_1 \mid x_1, s^j) \log \frac{P(y_1 \mid x_1 s^j)}{P(y_1 \mid s^j)} \quad (9)$$

where $\mu_1(\cdot)$ is the probability distribution on A. ($x_1 \in A$, $y_1 \in B$, $s^j \in S$). This capacity is identical with Wolfowitz's result (Wolfowitz, 1964, p. 34). The following sections include the proofs of the coding and the converse theorems.

2. THE CODING THEOREM

LEMMA 1. *For any real number D and any $j = 0, \pm 1, \pm 2, \cdots$ let*

$$A^j = \left\{ (X_n^j, Y_n^j) : \log \frac{P(Y_n^j \mid X_n^j)}{P(Y_n^j)} > nD \right\} \quad (10)$$

then there exist a code $[[e^{nR}], n, \lambda_n^j]$ for any positive integer R, such that

$$\lambda_n^j \leqq e^{n(R-D)} + \mathrm{P_r}\{(X_n^j, Y_n^j) \notin A^j\} \quad (11)$$

where $\mathrm{P_r}\{\ \}$ denotes the probability of the event $\{\ \}$.

Proof. Ash (1965, pp. 68–71), Wolfowitz (1964, pp. 97–98). It should be noted that $\lambda_n^j = \sum_{s^j} P(s^j) \lambda_{n,s}^j$ and that $P(Y_n^j \mid X_n^j) = \sum_{s^j} P(s^j) P(Y_n^j \mid X_n^j, s^j)$.

LEMMA 2. *Let*

$$i^j(n) = \frac{1}{n} \log \frac{P(Y_n^j \mid X_n^j)}{P(Y_n^j)} \quad (12a)$$

Then, for any D

$$\mathrm{P_r}\{i^j(n) < D\} \leqq e^{-n\rho(-D+G^j(\rho,n))}; \quad \rho > 0$$

Where

$$G^j(\rho, n) = -\frac{1}{\rho n} \log \sum_{\{X_n^j, Y_n^j\}} P(X_n^j, Y_n^j) \frac{P(Y_n^j \mid X_n^j)^{-\rho}}{P(Y_n^j)^{-\rho}} \quad (12b)$$

Proof. Let $\rho > 0$, then

$$P(I^j(n) < D) = \sum_{\{(X_n{}^j, Y_n{}^j): I^j(n) < D\}} P(X_n{}^j, Y_n{}^j)$$
$$\leqq \sum_{\{X_n{}^j, Y_n{}^j\}} e^{n\rho[D-I^j(n)]} P(X_n{}^j, Y_n{}^j) = e^{n\rho[D-G^j(\rho,n)]}$$

LEMMA 3. *Let*

$$G^j_{Z_p{}^j}(\rho, n) = -\frac{1}{\rho n} \log \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid Z_p^i) \frac{P(Y_n{}^j \mid X_n{}^j, Z_p{}^j)^{-\rho}}{P(Y_n{}^j \mid Y_p^{j-p})^{-\rho}}$$

where $Z_p{}^j$ *is the event* $\{X_p^{j-p}, Y_p^{j-p}\}$ *and where* $p = 1, 2, \cdots$ *and let* $0 < \rho < 1$, *then*

$$\frac{n}{n-1} G^{j_0-1}(\rho, n) \geqq \inf_K \inf_{p, Z_p{}^j} G^j_{Z_p{}^j}(\rho, n-1) \triangleq \min G^0(\rho, n-1) \tag{13}$$
$$p = 1, 2, \cdots \quad j = j_0 \pm K(n-1); \quad K = 0, 1, 2, 3$$

Proof. See Appendix.

LEMMA 4. *Let* $m > n$ $(m = 1, 2, 3 \cdots ; n = 1, 2, 3 \cdots)$. *Then*

$$\frac{m}{n} G^j_{Z_p{}^j}(\rho, m) \geqq G^{j+t}_{Z_p{}^{j+t}}(\rho, n); \qquad t = 0, 1, 2, \cdots (m-n) \tag{14}$$

Proof. See Appendix.

LEMMA 5. *Let* $n = lq$ $(l = 1, 2, 3 \cdots , q = 1, 2, 3 \cdots)$ *and let*

$$\min G^0(\rho, q) \triangleq \inf_K \inf_{p, Z_p{}^j} G^j_{Z_p{}^j}(\rho, q); \tag{15a}$$
$$j = j_0 \pm Kq; \quad K = 0, 1, 2, 3; \quad p = 1, 2 \cdots$$

then

$$\min G^0(\rho, lq) \geqq \min G^0(\rho, q) \tag{15b}$$

Proof. See Appendix.

LEMMA 6. *Let*

$$I^j_{Z_p{}^j}(n) = \frac{1}{n} \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) \log \frac{P(Y_n{}^j \mid X_n{}^j, Z_p{}^j)}{P(Y_n{}^j \mid Y_p^{j-p})} \tag{16a}$$

Let

$$m > n \ (m = 1, 2, 3 \cdots n = 1, 2, 3 \cdots)$$

Then

$$\frac{m}{n} I^{j}_{Z_p^{j}}(m) \geqq I^{j+t}_{Z_p^{j+t}}(n); \qquad t = 0, 1, 2, \cdots m - n. \tag{16b}$$

Proof. See Appendix.

LEMMA 7. *Let* $n = lq$ $(l = 1, 2, 3 \cdots ; q = 1, 2, 3)$, *and let*

$$\min I^{0}(n) \triangleq \inf_{K} \inf_{p,\, Z_p^{j}} I^{j}_{Z_p^{j}}(n);$$

$$j = j_0 \pm Kn; \quad K = 0, 1, 2, 3; \quad p = 1, 2 \cdots$$

Then

$$\min I^{0}(lq) \geqq \min I^{0}(q) \tag{17}$$

Proof. See Appendix

LEMMA 8. *The limit of* $\min I^{0}(n)$ *as* n *goes to infinity, always exists. Let this limit be denoted by:*

$$C^{0} \triangleq \lim_{n\to\infty} \min I^{0}(n) = \lim_{n\to\infty} \inf_{K} \inf_{p,\, Z_p^{j}} I^{j}_{Z_p^{j}}(n)$$
$$j = j_0 \pm Kn; \quad K = 0, 1, 2, 3 \cdots ; \quad p = 1, 2 \cdots \tag{18}$$

Proof. Follows from Lemma 6 and Lemma 7; See Appendix.

LEMMA 9. *The function* $G^{j}_{Z_p^{j}}(\rho, q)$ *is lower bounded uniformally with respect to all* Z_p^{j} *and* j, *by*

$$G^{j}_{Z_p^{j}}(\rho, q) \geqq I^{j}_{Z_p^{j}}(q) - M^{q}\left(\frac{4}{e}\right)^{2} \frac{\rho}{q(1-\rho)^{2}} \tag{19}$$

where M *is the number of letters in the output alphabet (the set* B.)

Proof. See Appendix.

Now, by Lemma 8, it is possible to choose an integer q such that for any $n > q$ and any positive number ϵ_1,

$$\min I^{j}_{Z_p^{j}}(q) > C^{0} - \epsilon_1 \tag{20}$$

By Lemma 9, it is possible to choose a number ρ_0, $(0 < \rho_0 < 1)$, such that for any positive number ϵ_2,

$$\min G^{0}(\rho_0, q) \geqq \min I^{0}(q) - \epsilon_2 \tag{21}$$

Let l be an integer such that for $n - 1 > q$

$$lq < n - 1 \leqq (l + 1)q$$

Then, by Lemma 4 and Lemma 5 we have,

$$\min G^0(\rho_0, n-1) \geqq \frac{lq}{n-1} \min G^0(\rho_0, q) \geqq \frac{l}{l+1} \min G^0(\rho_0, q) \tag{22}$$

Inserting Eqs. (21) and (20) into Eq. (22) yields

$$\min G^0(\rho_0, n-1) \geqq \frac{l}{l+1} [\min I^0(q) - \epsilon_2] \geqq \frac{l}{l+1} [C^0 - \epsilon_1 - \epsilon_2]$$

and therefore, for sufficiently large n (and hence for appropriately large l),

$$\min G^0(\rho_0, n-1) \geqq C^0 - 2(\epsilon_1 + \epsilon_2) \tag{23}$$

Inserting Eqs. (23) and (13) into Eq. (12) yields

$$\mathrm{P_r}\{i^j(n) < D\} \leqq \exp\{n\rho_0[D - C^0 + 2(\epsilon_1 + \epsilon_2)]\}; \qquad j = j_0 - 1 \tag{24}$$

for large enough n.
Now let

$$\epsilon = 4(\epsilon_1 + \epsilon_2)$$
$$D = C^0 - \epsilon$$

then, for large enough n,

$$\mathrm{P_r}\{i^j(n) < C^0 - \epsilon\} \leqq e^{-n\rho_0\epsilon/2}; \qquad j = j_0 - 1 \tag{25}$$

where ϵ is an arbitrary-small positive number.

Now C^0 is a function of the specific probability measure μ on A^1; Thus, let

$$\bar{C} = \inf_{j_0} \sup_{\mu} C^0 = \inf_{j_0} \sup_{\mu} \lim_{n\to\infty} \inf_{K} \inf_{P,\, Z_n{}^j} I^j_{Z_p{}^j}(n) \qquad j = j_0 \pm Kn; \quad K = 0, 1, 2, 3 \cdots \tag{26}$$

The insertion of Eqs. (26) and (25) into Eq. (11) yields, for any j and any large enough n,

$$\lambda_n{}^j \leqq e^{n(R-\bar{C}+\epsilon)} + e^{-n\rho_0\epsilon/2}$$

where ϵ is an arbitrary positive number.

Therefore, for large enough n,

$$\lambda_n{}^j \leqq e^{-E(R)\cdot n}; \quad E(R) > 0 \quad \text{for} \quad R < \bar{C}. \tag{27}$$

Now,

$$\sum_{\{Y_n{}^j\}} P(Y_n{}^j \mid Z_p{}^j, s^{j-p}) \log \frac{P(Y_n{}^j \mid Y_p^{j-p})}{P(Y_n{}^j \mid Z_p{}^j, s^{j-p})}$$

$$\leqq \sum_{\{Y_n{}^j\}} P(Y_n{}^j \mid Z_p{}^j, s^{j-p}) \left(\frac{P(Y_n{}^j \mid Y^{j-p})}{P(Y_n{}^j \mid Z_p{}^j, s^{j-p})} - 1 \right) = 0$$

Therefore

$$\begin{aligned} &-\frac{1}{n} \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) \log P(Y_n{}^j \mid Y_p^{j-p}) \\ &= -\frac{1}{n} \sum_{S^{j-p}} P(s^{j-p}) \sum_{\{Y_n{}^j\}} P(Y_n{}^j \mid Z_p{}^j, s^{j-p}) \log P(Y_n{}^j \mid Y_p^{j-p}) \\ &\qquad \leqq -\frac{1}{n} \sum_{S^{j-p}} P(s^{j-p}) \sum_{\{X_n{}^j, Y_n{}^j\}} \\ &\qquad\qquad \cdot P(X_n{}^j, Y_n{}^j \mid s^j) \log P(Y_n{}^j \mid s^j) \end{aligned} \tag{28}$$

where $s^j = \{s^{j-p}; Z_p{}^j\}$

Now,

$$\begin{aligned} &\log P(Y_n{}^j \mid Z_p{}^j, X_n{}^j) = \log \sum_{S^{j-p}} P(s^{j-p}) P(Y_n{}^j \mid X_n{}^j, s^j) \\ &\geqq \log P(s^{j-p}) + \log P(Y_n{}^j \mid X_n{}^j, s^j); \quad s^{j-p} \text{ is any element of } S^{j-p} \end{aligned} \tag{29}$$

The insertion of Eqs. (28) and (29) into Eq. (16a) yields:

$$\begin{aligned} I_{Z_p{}^j}{}^j(n) &\geqq \frac{1}{n} \sum_{\{s^{j-p}\}_\epsilon} P(s^{j-p}) \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid s^j) \log \frac{P(Y_n{}^j \mid X_n{}^j, s^j)}{P(Y_n{}^j \mid s^j)} \\ &\quad + \frac{1}{n} \sum_{\{s^{j-p}\}_\epsilon} P(s^{j-p}) \log P(s^{j-p}) = \sum_{\{s^{j-p}\}_\epsilon} P(s^{j-p}) I_s{}^j(n) \\ &\quad + \frac{1}{n} \sum_{\{s^{j-p}\}_\epsilon} P(s^{j-p}) \log P(s^{j-p}) \end{aligned} \tag{30}$$

where

$$1)\ I_s{}^j(n) = \frac{1}{n} \sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid s^j) \log \frac{P(Y_n{}^j \mid X_n{}^j, s^j)}{P(Y_n{}^j \mid s^j)}$$

2) The set $\{s^{j-p}\}_\epsilon$ is any finite subset of the countable set S^{j-p} such that $P(\{s^{j-p}\}_\epsilon) \geqq 1 - \epsilon$, where ϵ is an arbitrary positive number. Now

$$-\sum_{\{s^{j-p}\}_\epsilon} P(s^{j-p}) \log P(s^{j-p})$$

$$\leqq (\log K)(1 - \epsilon) - (1 - \epsilon) \log (1 - \epsilon) \leqq \log 2K$$

where K is the number of states in $\{s^{j-p}\}_\epsilon$.
Thus,

$$I^j_{Z^j_p} \geqq (1 - \epsilon) \inf_{s^j \epsilon S^j} I_s{}^j(n) - \frac{\log 2K}{n} \tag{31}$$

and for large enough n,

$$I^j_{Z^j_p} \geqq \inf_{s^j \epsilon S^j} I_s{}^j(n) - \epsilon_1$$

where ϵ_1 is an arbitrary positive number.

Therefore by Eqs. (27) and (31) and for large enough n

$$\lambda_n{}^j \leqq e^{-E(R)n} \tag{32}$$

where $E(R)$ is positive for any $R < C_s$ and where

$$C_s = \inf_{j_0} \sup_{\mu} \lim_{n \to \infty} \inf_{K} \inf_{s^j} I_s{}^j(n) \tag{33}$$

This proves section (b), of the main theorem.

Now, given some arbitrary j_0, let μ be a probability measure on A^1 that maximizes C in Eq. (3) for this specific j_0; Then

$$P_r(i^{j_0-1}(n) < C - \epsilon) = P_r(i^{j_0-1}(n) < C - \epsilon; I_s{}^j(n) \geqq C)$$

$$+ P_r(i^{j_0-1}(n) < C - \epsilon; I_s{}^j(n) < C)$$

$$\leqq P_r(i^{j_0-1}(n) < I_s{}^j(n) - \epsilon) + P_r(I_s{}^j(n) < C)$$

for any $j = j_0 - 1 \pm Kn$; $K = 0, 1, 2, 3 \cdots$

Now, by Eqs. (18) and (31), one can choose j such that for large enough n,

$$\inf_{s^j} I_s{}^j(n) < C^0 + \frac{\epsilon}{2}$$

Therefore,

$$P_r(i^{j_0-1}(n) < C - \epsilon) \leqq P_r\{i_{(n)}^{j_0-1} < C^0 - \frac{\epsilon}{2}\}$$
$$+ \sup_K P_r\{I_s^j(n) < C\}; j = j_0 - 1 \pm Kn; K = 0, 1, 2, 3 \cdots$$

Thus, by Eqs. (1), (2), (3) and (25)

$$\lim_{n\to\infty} P_r(i^j(n) < C - \epsilon) = 0 \tag{34}$$

for any j and any positive ϵ.

Now, let $D = C - \epsilon$; then, by Eqs. (11) and (34)

$$\lim_{n\to\infty} \lambda_n{}^j = 0; \quad \text{for} \quad R < C$$

This proves section (a) of the coding theorem.

Eqs. (33) holds for any probability distribution on S^j, and therefore also for the case where $P(s^j) > 0$ for all $s^j \in S^j$.
Thus

$$\lambda_n{}^j = \sum_{s^j} P(s^j)\lambda_{n,s}^j \geqq P(s^j)\lambda_{n,s}^j \text{; any } s^j \in S^j$$

$$\therefore \lambda_{n,s}^j \leqq \frac{1}{P(s^j)} \lambda_{n,s}^j \leqq \frac{1}{P(s^j)} e^{-E(R)\cdot n} = K(s^j)e^{-E(R)\cdot n}$$

where $K(S^j)$ is finite and is independent of R and n.

This proves section (c) of the main theory.

3. THE CONVERSE THEOREM

Let $\lambda_{n,s}^j$ be the probability of error that is associated with a code: $X_{n,1}^j$, $X_{n,2}^j \cdots X_{n,e^{nR}}^j$, given that the channel is at the state s^j. Let us define a probability distribution on the set $\{X_n{}^j\}$ such that

$$\begin{aligned} P(X_n{}^j) &= e^{-nR}; \qquad X_n{}^j = X_{n,i}^j \quad i = 1 \cdots e^{nR} \\ P(X_n{}^j) &= 0; \qquad\quad X_n{}^j \neq X_{n,i}^j \quad i = 1 \cdots e^{nR} \end{aligned} \tag{35}$$

Then, by Fano's inequality (Ash, 1965; Wolfowitz, 1964),

$$\begin{aligned} &-\sum_{\{X_n{}^j, Y_n{}^j\}} P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) \log P(X_n{}^j \mid Y_n{}^j, s^j) \\ &\leqq \lambda_{n,s}^j \log (e^{nR} - 1) - \lambda_{n,s}^j \log \lambda_{n,s}^j - (1 - \lambda_{n,s}^j) \log (1 - \lambda_{n,s}^j) \end{aligned} \tag{36}$$

Now, by Eq. (35)

$$P(X_n^j) = P(X_n^j \mid s^j) = e^{-nR}; \qquad X_n^j = X_{n,i}^j \quad i = 1 \cdots e^{nR}$$
$$= 0; \qquad X_n^j \neq X_{n,i}^j \quad i = 1 \cdots e^{nR}$$

Therefore,

$$-\sum_{\{X_n^j, Y_n^j\}} P(X_n^j, Y_n^j \mid s^j) \log P(X_n^j \mid s^j) = nR \tag{37}$$

Now, by Eq. (37) and (36) we get:

$$\lambda_{n,s}^j \log (e^{nR} - 1) - \lambda_{n,s}^j \log \lambda_{n,s}^j - (1 - \lambda_{n,s}^j) \log (1 - \lambda_{n,s}^j)$$
$$\geqq nR - \sum_{\{X_n^j, Y_n^j\}} P(X_n^j, Y_n^j \mid s^j) \log \frac{P(Y_n^j \mid X_n^j, s^j)}{P(Y_n^j \mid s^j)} \tag{38}$$

Thus,

$$\lambda_{n,s}^j - \frac{\lambda_{n,s}^j \log \lambda_{n,s}^j + (1 - \lambda_{n,s}^j) \log (1 - \lambda_{n,s}^i)}{nR}$$
$$\geqq \frac{R - \frac{1}{n} \sum_{\{X_n^j, Y_n^j\}} P(X_n^j, Y_n^j \mid s^j) \log \frac{P(Y_n^j \mid X_n^j, s^j)}{P(Y_n^j \mid s^j)}}{R} \tag{39}$$

Thus, by Eq. (1)

$$\lambda_{n,s}^j \geqq \frac{R - I_s^j(n)}{R} - \frac{\log 2}{n \cdot R} \tag{40}$$

(Since $-P \log P - (1 - P) \log (1 - P) \leqq \log 2; 0 \leqq P \leqq 1$)

Let us denote by $F_n^j(\epsilon)$ the set:

$$F_n^j(\epsilon) = \{s^j : I_s^j(n) \leqq C + \epsilon\} \tag{41}$$

It follows from Eqs. (1) (2) and (3) that for any $\epsilon > 0$ there exists a positive number η such that for any j_0, any probability measure μ on A^1 and any n_0, there exists some number $n > n_0$ for which

$$\Pr \{s^j \in F_n^j(\epsilon)\} \geqq \eta > 0 \tag{42}$$

for some $j = j_0 \pm Kn$; $K = 0, 1, 2, 3 \cdots$

Let us pick the one μ for which

$$P(X_n^j) = e^{-nR}; \qquad X_n^j = X_{n,i}^j \quad i = 1, 2, 3 \cdots e^{nR};$$
$$P(X_n^j) = 0; \qquad X_n^j \neq X_{n,i}^j \quad j = j_0 \pm Kn$$

where $\{X_{n,1}^{j} \cdots X_{n,i} \cdots X_{n,e^{nR}}^{j}\}$ is the best code of length n to be used at the jth instant (i.e. the one that minimizes $\lambda_n{}^j$)

Thus, μ is given by:

$$\cdots P(X_n^{j_0-n})P(X_n^{j_0})P(X^{j_0+n}) \cdots .$$

Then, by Eq. (40)

$$\lambda_{n,s}^{j} \geqq \frac{R - I_s{}^j(n)}{R} - \frac{\log 2}{nR}\,; s^j \in F_n{}^j(\epsilon) \tag{43}$$

for any $j = j_0 \pm Kn$; $K = 0, 1, 2, 3 \cdots$

Thus, by Eqs. (41) (42) and (43)

$$\lambda_n{}^j \geqq \eta \left[\frac{R - (C + \epsilon)}{R} - \frac{\log 2}{nR}\right] \quad \text{for some} \quad j = j_0 \pm Kn$$

since

$$\lambda_n{}^j = \sum_{s^j} P(s^j)\lambda_{n,s}^{j} > \sum_{s^j \epsilon F^j{}_n(\epsilon)} P(s^j)\lambda_{n,s}^{j}$$

Thus, for large enough n

$$\lambda_n{}^j \geqq \eta \frac{R - (C + 2\epsilon)}{R}\,; \quad \text{for some} \quad j_0 \pm Kn$$

where ϵ is an arbitrary positive number.

Therefore, for any $R > C$

$$\lambda_n{}^j \geqq \lambda > 0 \quad \text{for some} \quad j.$$

This proves section (a) of the converse theorem.

Now, it follows from Eq. (4) that for any n_0 there exist some $n > n_0$, some j and some state s^j such that for any μ

$$I_s{}^j(n) \leqq C_s + \epsilon_1 \tag{44}$$

where ϵ_1 is an arbitrary small positive number.

Inserting Eq. (44) into Eq. (40) yields for some $n > n_0$, some j and some $s_p{}^j$

$$\lambda_{n,s}^{j} > \frac{R - C - \epsilon_1}{R} - \frac{\log 2}{nR} \tag{45}$$

Thus, for large enough n_0 there exists some $n > n_0$ such that

$$\lambda^j_{n,s} > \frac{R - C - \epsilon_1}{R} \tag{46}$$

for some j and some s^j, where ϵ_1 is an arbitrary small positive number.

This proves section (b) of the converse theorem.

4. APPENDIX

Proof of Lemma 3

$$G^{j_0-1}(\rho, n) = -\frac{1}{\rho n} \log \sum_{\{X_n^{j_0-1}, Y_n^{j_0-1}\}} \cdot P(X_1^{j_0-1}, Y_1^{j_0-1}) P(X_{n-1}^{j_0}, Y_{n-1}^{j_0} \mid X_1^{j_0-1}, Y_1^{j_0-1}) \cdot \left| \frac{P(Y_1^{j_0-1} \mid X_1^{j_0-1})}{p(Y_1^{j_0-1})} \right|^{-\rho} \left| \frac{P(Y_{n-1}^{j_0} \mid X_{n-1}^{j_0}, X_1^{j_0-1}, Y_1^{j_0-1})}{P(Y_{n-1}^{j_0} \mid Y_1^{j_0-1})} \right|^{-\rho}$$

Let us denote by $Z_p{}^j$ the event $\{X_p^{j-p}, Y_p^{j-p}\}$.

Then,

$$G^{j_0-1}(\rho, n) = -\frac{1}{\rho n} \log \sum_{\{X_1^{j_0-1}, Y_1^{j_0-1}\}} P(X_1^{j_0-1}, Y_1^{j_0-1}) \left| \frac{P(Y_1^{j_0-1} X_1^{j_0-1})}{P(Y_1^{j_0-1})} \right|^{-\rho} \cdot \sum_{\{X_{n-1}^{j_0}, Y_{n-1}^{j_0}\}} P(X_{n-1}^{j_0}, Y_{n-1}^{j_0} \mid Z_1^{j_0}) \cdot \frac{P(Y_{n-1}^{j_0} \mid X_{n-1}^{j_0}, Z_1^{j_0})}{P(Y_{n-1}^{j_0} \mid Y_1^{j_0-1})}$$

Thus,

$$G_0^{j-1}(\rho, n) \geqq -\frac{1}{\rho n} \log \sum_{\{X_1^{j_0-1}, Y_1^{j_0-1}\}} P(X_1^{j_0-1}, Y_1^{j_0-1}) \frac{P(Y_1^{j_0-1} \mid X_1^{j_0-1})^{-\rho}}{P(Y_1^{j_0-1})^{-\rho}}$$

$$-\frac{1}{\rho n} \log \max_{z_1^{j_0}} \sum_{\{X_{n-1}^{j_0}, Y_{n-1}^{j_0}\}} P(X_{n-1}^{j_0}, Y_{n-1}^{j_0} \mid Z_1^{j_0}) \frac{P(Y_{n-1}^{j_0} \mid X_{n-1}^{j_0}, Z_1^{j_0})^{-\rho}}{P(Y_{n-1}^{j_0} \mid Y_1^{j_0-1})^{-\rho}}$$

$$= \frac{1}{n} G^{j_0-1}(\rho, 1) + \frac{n-1}{n} \min_{z_1^{j_0}} G^{j_0}_{z_1^{j_0}}(\rho, n-1)$$

$$\geqq \frac{1}{n} G^{j_0-1}(\rho, 1) + \frac{n-1}{n} \inf_{K,p} \min_{Z_p{}^j} G^j_{Z_p{}^j}(\rho, n-1)$$

$$j = j_0 \pm K.n \quad K = 0, 1, 2, 3 \cdots ; p = 1, 2 \cdots$$

Now,

$$G^{j_0-1}(\rho, 1) = -\frac{1}{\rho}\log \sum_{\{X^{j_0-1}, Y^{j_0-1}\}} P(X^{j_0-1}, Y^{j_0-1}) \left[\frac{P(Y_1^{j_0-1} \mid X_1^{j_0-1})}{P(Y_1^{j_0-1})}\right]^{-\rho}$$

$$\geqq -\frac{1}{\rho} \sum_{\{X_1^{j_0-1}, Y_1^{j_0-1}\}} P(X_1^{j_0-1}, Y_1^{j_0-1}) \log \left[\frac{P(Y_1^{j_0-1} \mid X_1^{j_0-1})}{P(Y^{j_0-1})}\right]^{-\rho} =$$

$$\geqq 0$$

Thus,

$$\frac{n}{n-1} G^{j_0-1}(\rho, n) \geqq \inf_{K, p} \min_{Z_p{}^j} G^j_{Z^j_p}(\rho, n-1) \triangleq \min G^0(\rho, n-1)$$

$$j = j_0 \pm Kn; \quad K = 0, 1, 2, 3 \cdots ; p = 1, 2 \cdots$$

Proof of Lemma 4

$$P(X_m{}^j, Y_m{}^j \mid Z_p{}^j) = P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) P(X_{m-n}^{j+n}, Y_{m-n}^{j+n} \mid X_n{}^j, Y_n{}^j, Z_p{}^j)$$

$$= P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) P(X_{m-n}^{j+n}, Y_{m-n}^{j+n} \mid Z_{p+n}^{j+n})$$

Thus,

$$mG^j_{Z^j_p}(\rho, m) = -\frac{1}{\rho}\log \left[\sum_{X_n{}^j} P(X_n{}^j, Y_n{}^j \mid Z_p{}^j) \frac{P(Y_n{}^j \mid Y_n{}^j, Z_p{}^j)^{-\rho}}{P(Y_n{}^j \mid Y_p^{j-p})^{-\rho}}\right.$$

$$\left. \sum_{\{X_{m-n}^{j+n}, Y_{m-n}^{j+n}\}} P(X_{m-n}^{j+n}, Y_{m-n}^{j+n} \mid Y_{p+n}^{j+n}) \times \frac{P(Y_{m-n}^{j+n} \mid X_{m-n}^{j+n}, Z_{p+n}^{j+n})^{-\rho}}{P(Y_{m-n}^{j+n} \mid Y_{p+n}^{j-p})^{-\rho}}\right]$$

Therefore it follows that,

$$mG^j_{Z^j_p}(\rho, m) \geqq nG^j_{Z^j_p}(\rho, n) + (m-n) \inf_{j, p} \min_{Z_p{}^j} G^j_{Z^j_p}(\rho, m-n)$$

Now, similar to the proof of Lemma 3 it can be shown that

$$G^j_{Z^j_p}(\rho, m-n) \geqq 0; \quad \text{any} \quad j \quad \text{and any} \quad Z_p{}^j.$$

Thus, for any j and any $Z_p{}^j$

$$mG^j_{Z^j_p}(\rho, m) \geqq nG^j_{Z^j_p}(\rho, n)$$

In a similar way it can be shown that

$$mG^j_{Z^j_p}(\rho, m) \geqq nG^{j+t}_{Z^{j+t}_p}(\rho, n); \quad t = 0, 1, 2 \cdots m-n.$$

PROOF OF LEMMA 5

$$P(X_{lq}^{i}, Y_{lq}^{i} \mid Z_{p}^{j}) = \prod_{s=0}^{l-1} P(X_{q}^{j+sq}, Y_{q}^{j+sq} \mid Z_{p+sq}^{j+sq})$$

Thus, similar to the proof of Lemma 3, it follows that

$$G_{Z_{p}^{j}}^{j}(\rho, lq) \geqq \frac{1}{\rho l q} lq \inf_{K, p} \min_{Z_{p}^{j}} G_{Z_{p}^{j}}^{j}(\rho, q)$$

Hence

$$G_{Z_{p}^{j}}^{j}(\rho, lq) \geqq \min G^{0}(\rho, q)$$

for any $j = 0, \pm 1, \pm 2 \cdots$
Therefore

$$\min G^{0}(\rho, lq) \geqq \min G^{0}(\rho, q)$$

PROOF OF LEMMA 6

$$mI_{Z_{p}^{j}}^{j}(m) = \sum_{\{X_{n}^{j}, Y_{n}^{j}\}} P(X_{n}^{j}, Y_{n}^{j} \mid Z_{p}^{j}) \left\{ \log \frac{P(Y_{n}^{j} \mid X_{n}^{j}, Z_{p}^{j})}{P(Y_{n}^{j} \mid Y_{p}^{j-p})} \right.$$

$$\left. + \sum_{\{X_{m+n}^{j+n}, Y_{m-n}^{j+n}\}} P(X_{m-n}^{j+n}, Y_{m-n}^{j+n} \mid Z_{p+n}^{j+n}) \log \frac{P(Y_{m-n}^{j+n} \mid X_{m-n}^{j+n}, Z_{p+n}^{j+n})}{P(Y_{m-n}^{j+n} \mid Y_{p+n}^{j-p})} \right\}$$

Now, it can be shown that

$$\sum_{\{X_{m-n}^{j+n}, Y_{m-n}^{j+n}\}} P(X_{m-n}^{j+n}, Y_{m-n}^{j+n} \mid Z_{n+p}^{j+n}) \log \frac{P(Y_{m-n}^{j+n} \mid X_{m-n}^{j+n}, Z_{n+p}^{j+n})}{P(Y_{m-n}^{j+n} \mid Y_{p+n}^{j-p})} \geqq 0$$

Thus,

$$mI_{Z_{p}^{j}}^{j}(m) \geqq nI_{Z_{p}^{j}}^{j}(n)$$

In a similar way it can be shown that

$$mI_{Z_{p}^{j}}^{j}(m) \geqq nI_{Z_{q}^{j+t}}^{j+t}(n); \quad t = 0, 1, 2, \cdots m - n.$$

PROOF OF LEMMA 7

Similar to the proof of Lemma 5.

PROOF OF LEMMA 8

Let

$$a = \lim_{n \to \infty} \sup \{\min I^0(n)\}$$

Now

$$0 \leqq I^j_{Z_p{}^j}(n) \leqq -\frac{1}{n} \sum_{\{Y_n{}^j\}} P(Y_n{}^j \mid Z_p{}^j) \log P(Y_n{}^j \mid Z_p{}^j) \leqq \log M$$

where M is the number of letters in the output alphabet, B.

Thus

$$0 \leqq a \leqq \log M < \infty$$

Let ϵ be given arbitrarily and let the subscript q be chosen such that

$$\min I^0(q) > a - \epsilon$$

For any $n > q$ we determine l ($l = 1, 2, 3 \cdots$) such that $lq \leqq n \leqq (l+1)q$. By Lemma 6 and Lemma 7 we have

$$\min I^0(n) \geqq \frac{lq}{n} \min I^0(lq) \geqq \frac{l}{l+1} \min I^0(q)$$

Thus

$$\min I^0(n) > \frac{l}{l+1} (a - \epsilon)$$

Therefore, for sufficiently large n (and hence for appropiately large l):

$$a + \epsilon \geqq \min I^0(n) > a - 2\epsilon$$

But since ϵ is arbitrarily small

$$\lim_{n \to \infty} \min I^0(n) = a = C^0.$$

PROOF OF LEMMA 9

Since $\log X \leqq X - 1$ for any X, we have

$$1 - qpG^j_{Z_p{}^j}(\rho, q) \leqq \sum_{\{X_q{}^j, Y_q{}^j\}} P(X_q{}^j, Y_q{}^j \mid Z_p{}^j) \cdot \frac{P(Y_q{}^j \mid X_q{}^j, Z_p{}^j)^{-\rho}}{P(Y_q{}^j \mid Y_p{}^{j-p})^{-\rho}} \triangleq f_{Z_p{}^j}(\rho, q) - 1 \tag{A-1}$$

Now,

$$f_{Z_p^i}(0, q) = 1 \tag{A-2}$$

$$\left.\frac{\partial f_{Z_p^i}(\rho, q)}{\partial \rho}\right|_{\rho=0} = -qI_{Z_p^i}^i(q) \tag{A-3}$$

Also

$$\begin{aligned}
\frac{\partial^2 f_{Z_p^i}(\rho, q)}{\partial \rho^2} &= \sum_{\{X_q^j, Y_q^j\}} P(X_q^j, Y_q^j \mid Z_q^j) \\
&\quad \cdot \frac{P(Y_q^j \mid X_p^j, Z_p^j)^{-\rho}}{P(Y_q^j \mid Y_p^{j-p})^{-\rho}} \log^2 \frac{P(Y_q^j \mid X_q^j, Z_p^j)}{P(Y_q^j \mid Y_p^{j-p})} \\
&\leqq \sum_{\{X_q^j, Y_q^j\}} P(X_q^j, Y_q^j \mid Z_p^j) \frac{P(Y_q^j \mid X_q^j, Z_p^j)^{-\rho}}{P(Y_q^j \mid Y_p^{j-p})^{-\rho}} \\
&\quad \cdot [4 \log^2 P(Y_q^j \mid X_q^j, Z_p^j) + 4 \log^2 P(Y_q^j \mid Y_p^{j-p})] \\
&\leqq 4 \sum_{\{X_q^j, Y_q^j\}} P(X_q^j \mid Z_p^j) P(Y_q^j \mid X_q^j, Z_p^j)^{1-\rho} \\
&\quad \log^2 P(Y_q^j \mid X_q^j, Z_p^j) \\
&\quad + 4 \sum_{\{X_q^j, Y_q^j\}} P(X_q^j \mid Z_p^j) P(Y_q^j \mid X_q^j, Z_p^j)^{1-\rho} \\
&\quad \log^2 P(Y_q^j \mid Y_p^{j-p})
\end{aligned}$$

Now, for $0 \leqq \rho < 1$ we get by convexity arguments:

$$\begin{aligned}
&\sum_{\{X_q^j\}} P(X_q^j \mid Z_p^j) P(Y_q^j \mid X_q^j, Z_p^j)^{1-\rho} \\
&\qquad \leqq \left[\sum_{\{X_q^j\}} P(X_q^j \mid Z_p^j) P(Y_q^j \mid X_q^j, \ Z_p^j)\right]^{1-\rho} = P(Y_q^j \mid Z_q^j)^{1-\rho}
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial^2 f_{Z_p^i}(\rho, q)}{\partial \rho^2} &\leqq \frac{4}{(1-\rho)^2} \sum_{\{X_q^j\}} P(X_q^j \mid Z_p^j) \sum_{\{Y_q^j\}} P(Y_q^j \mid X_q^j, Z_p^j)^{1-\rho} \\
&\qquad \log^2 P(Y_q^j \mid X_q^j, Z_p^j)^{1-\rho} \\
&\quad + \frac{4}{(1-\rho)^2} \sum_{\{X_q^j\}} P(Y_q^j \mid Z_p^j)^{1-\rho} \log^2 P(Y_q^j \mid Y_p^{j-p})
\end{aligned}$$

Now, by convexity arguments for $\log^2 X$, $(X < 1)$:

$$\log^2 P(Y_q^j \mid Y_p^{j-p}) = \log^2 \sum_{\{X_p^{j-p}\}} P(Y_q^j \mid Z_p^j) P(X_p^{j-p} \mid Y_p^{j-p})$$

$$\leqq \sum_{\{X_p^{j-p}\}} P(X_p^{j-p} \mid Y_p^{j-p}) \log^2 P(Y_q^j \mid Z_p^j)$$

Furthermore, since for any $0 \leqq X < 1$, $X \log^2 X \leqq 4e^{-2}$, it follows that:

$$\frac{\partial^2 f_{Z_p^j}(\rho, q)}{\partial \rho^2} \leqq \frac{16e^{-2}}{(1-\rho)^2} \left[\sum_{\{Y_q^j\}} 1 + \sum_{\{Y_q^j\}} 1 \right] = \frac{32e^{-q}}{(1-\rho)^2} M^q \qquad \text{(A-4)}$$

where M is the number of letters in the output alphabet B.

By Taylor series expansion theorem we have

$$f_{Z_p^j}(\rho, q) = f_{Z_p^j}(0, q) + \rho \left. \frac{\partial f_{Z_p^j}(\rho, q)}{\partial \rho} \right|_{\rho=0} + \frac{\rho^2}{2} \left. \frac{\partial^2 f_{Z_p^j}(\rho, q)}{\partial \rho^2} \right|_{\theta\rho} \qquad \text{(A-5)}$$

where $0 \leqq \theta \leqq 1$.

Thus, by Eqs. (A-2), (A-3) and (A-4):

$$f_{Z_p^j}(\rho, q) \leqq 1 - \rho q I_{Z_p^j}^j(q) + \frac{\rho^2}{(1-\rho)^2} 16e^{-2} M^q. \qquad \text{(A-6)}$$

The insertion of Eq. (A-6) into Eq. (A-1) yields:

$$G_{Z_p^j}^j(\rho, p) \geqq I_{Z_p^j}^j(q) - M^q \left(\frac{4}{e}\right)^2 \frac{q(1-\rho)^2}{\rho}$$

ACKNOWLEDGMENT

This research was conducted under the auspices of the Scientific Department, Israel Ministry of Defence.

RECEIVED: April 25, 1968; revised January 29, 1969

REFERENCES

ASH, R. B. (1965), "Information Theory." Wiley (Interscience), New York.

WOLFOWITZ, J. (1964), "Coding Theorems of Information—Theory" (second Edition). Springer-Verlag, Berlin.

YUDKIN, H. L. (March 1967), On the exponential error bound and capacity for finite state channels. M.I.T. Lincoln Laboratory.